L
150
.58
V.6
no.3

Experiments with Reflective Teaching

By

ERNEST E. BAYLES
School of Education, University of Kansas

KANSAS STUDIES IN EDUCATION

UNIVERSITY OF KANSAS PUBLICATIONS
SCHOOL OF EDUCATION—LAWRENCE KANSAS

Vol. 6 APRIL, 1956 No. 3

UNIVERSITY OF MICHIGAN
GENERAL LIBRARY

FOREWORD

Educators interested in more effective ways of instructing youth will find the present issue of *Kansas Studies* unusually rewarding. Included are six experimental studies all of which attempt to examine a type of instruction known as "reflective teaching." Testing this method of instruction by different investigators in different communities and at different grade levels provides the reader with a breadth of analysis that is missing from many experimental reports.

Professor Bayles has been interested in "reflective teaching" for some time, and he has made a definite contribution to teaching theory by directing the series of studies reported herein. We are happy to present this report as our final issue of *Kansas Studies* for this school year, for we know it will help many to enrich their own teaching procedures.

KENNETH E. ANDERSON
OSCAR M. HAUGH

Table of Contents

	PAGE
INTRODUCTION	1
A STUDY IN AMERICAN GOVERNMENT	3
THE DROLL STUDY	3
THREE STUDIES IN AMERICAN HISTORY	7
THE AVERY STUDY	9
THE JOHNSTON STUDY	10
THE SAILER STUDY	12
TWO STUDIES IN UPPER ELEMENTARY GRADES	15
THE TREFZ STUDY	15
THE READER STUDY	23
GENERAL COMMENT	29

THE UNIVERSITY OF KANSAS SCHOOL OF EDUCATION

DEAN, SCHOOL OF EDUCATION ..... Kenneth E. Anderson

EDITOR, *Kansas Studies in Education* ..... Oscar M. Haugh

L
150
.58
V.6
No.3

Experiments with Reflective Teaching

ERNEST E. BAYLES
School of Education, University of Kansas

INTRODUCTION

In the course of educational history, and particularly within the past half-century, there has been much experimentation with teaching procedures. But there seems to have been very little dealing with genuinely reflective teaching. The term, *reflective teaching,* is herein taken to mean conducting classes in such manner as to promote reflective-scientific studies of student problems; studies which achieve the solid substance of Dewey's Complete Act of Thought or of modern scientific investigation.

It is the purpose of this report to present the outcomes of a number of experimental investigations involving serious and rather competent attempts to employ reflective teaching. It is not claimed that the employed procedures were, in all cases or even in any case, fully reflective. In fact, since all but the Trefz and Reader studies represented only the first time that the teacher had attempted to teach in this way, it would be an absurdity to assume that the procedures actually followed were undeviatingly reflective. As the Trefz and Reader studies will show, it appears that a teacher can gain competence in such teaching as she or he gains experience in it. The change-over is gradual rather than abrupt. Much is learned and achieved the first year, but each subsequent year provides an increment and a teacher is likely to be doing a much more thorough job the sixth year than the first.

Nor is it claimed that the measurement of results was, in all cases or even in any case, all that it should have been. All investigations were conducted with classes and students assigned to the teachers in the course of normal enrollments. The tests used as measuring instruments were those available to teachers in small and medium-size school systems of Kansas and Missouri. All studies were, in whole or in part, performed in connection with master's theses, prepared at the University of Kansas under the writer's advisement and hitherto unpublished. The studies are scattered over the period from 1940 to 1953, most of them in the early part.

Before describing the studies, however, a brief discussion of reflective teaching is in order. Classroom procedure follows the general form

of reflective-scientific study. The class is first maneuvered into a problem, an I-don't-know situation. Unless a group of students is faced with a question whose answer is to them unknown, no reflection is for them possible. It makes no difference who else knows the answer, the teacher or the whole outside world; if the answer to a given question is unknown to the members of the group involved, for them the question is "food for thought." In the words of one of the investigators on whom we shall report,

> In most cases the children come to us with the opinion that the teacher's word is final; she knows the answers. I found them amazed when, after we had discussed a problem, they turned to me for the final answer and I said, "I don't know." A problem had arisen when we were discussing what plants we need and one little girl mentioned flowers. A few scoffed at the idea, but she defended herself by saying, "You have to have prettiness and flowers are pretty." Naturally, the question was whether we had to have pretty things in order to live. After some discussion, they turned to me for the answer but were told, "I don't know." They paused, then asked how they could find out. The problem had been raised and we were ready and wanting to find a solution.

Thus, the first step in reflective teaching is problem raising. Then follows problem solving. Continuing the above quotation,

> Of course, we had first to decide what we meant by living. They finally reached the conclusion that we might "exist" without pretty things, but we couldn't really "live" without them. They had solved the problem to their own satisfaction, and they had learned much about plants and about living conditions in the process.

A problem-solving process may be short and simple, as it may seem in the case just described. Others may be long and involved. But, whether short, long, or in-between, it is reflective teaching only if the class members themselves are actively working and thinking the matter through to a conclusion. Of course, the teacher will give help; at times only a little, at times perhaps a great deal. Be such help little or much, however, the thought-line is that of the pupils and the conclusion is theirs. In fact, each pupil is encouraged—in extreme cases, perhaps, almost impelled—to come to his own conclusions, even though they may differ from the ones which represent majority agreement. In stating a conclusion, the teacher should persistently use the form of expression, "In light of our study, the conclusion seems to be that . . ." This provides an open way for later modification and for differences in personal conviction among the group members but, at the same time, provides a tangible and concise outcome—something to show for the effort.

The problem-solving aspect of reflective-scientific study takes the general pattern of formulating and testing hypotheses. Guesses are made as to what the solution might be; then deductions are drawn from the various guesses. Available data are studied in light of the various sets of deductions and the guess (or hypothesis) whose deductive pattern best fits the data is taken to be best. Further tailoring may perhaps be in order, to secure a better fit.

Next, deductions are made, looking toward data not already available, and steps are taken *to secure or to produce* those data, the latter representing experimentation. If such data turn out to be as predicted, the final step in scientific-reflective study has been taken and the guess or hypothesis responsible for such outcome is taken as the solution to the problem, or the conclusion. Such a conclusion, of course, although subsequently serving as the basis for designing whatever action or behavior may be within its domain, retains its tentative or hypothetical status; the conclusion to be held as long as it continues to satisfy the foregoing conditions but to be rejected as soon as it fails to do so.

Reflective teaching, then, is the name which we give to classroom procedures which follow the general form just indicated. A question is raised whose answer is unknown, study follows looking toward an answer, and finding the answer marks the culmination or conclusion of the study. The teacher is not giving silver-platter handouts. She or he is acting essentially as chairman of a group of investigators who are following modern scientific methods in carrying out their investigations. The classes, on which the following reports are made, were all conducted, in as far as each teacher was able under the circumstances to do so, in the foregoing manner.

A Study in American Government

The Droll Study

In 1940, Helen Andres Droll[1] reported an experiment with a high-school class in United States government composed of about twenty students from grades eleven and twelve, located in a small town in western Kansas. The class was taught during a spring semester. Two intelligence examinations showed the students to be not far from average; the median IQ on one test was 101 and on the other 102. In

[1] Helen Andres Droll. "A Study of a Semester's Work in Government at the High School Level, Presented in Accordance with the Dewey-Bode Point of View." Unpublished Master's thesis, University of Kansas, 1940.

fact, when compared with the national average of high-school juniors and seniors which is reported by the U.S. Office of Education to be about 107, this class would appear to have been a bit below average.[2]

The class was somewhat unusual in that a large portion was of foreign-born parentage. German or French was spoken in most of the homes. This situation might have established somewhat of a language handicap whereas, on the other hand, because of European parental background it might have established an unusual interest in problems relating to citizenship. The intelligence examinations were given, not to members of this class by themselves, but in connection with the regular high-school testing program and the scores were obtained from the regular file.

Five units were covered during the semester. They were as follows:

Unit I: Our dependence on others and how group life has grown in meaning and obligations
Unit II: What is government?
Unit III: How the will of the people is expressed
Unit IV: What is citizenship?
Unit V: It is said that the United States is in position to exert the moral leadership of the world today. How can we best exercise that moral leadership?

No single textbook was followed. Information was gained from whatever sources were available—a not very voluminous school library, free materials obtained from state and federal governmental and other sources, visits to local points of interest which were related to the studies, and talks by a few local citizens. Class meetings were discussional in nature, reflective in general form, and classroom examinations were given occasionally, both essay and objective types. The essay-type examination items tended to be interpretational whereas some of the objective-type items were distinctly factual.

Mrs. Droll was a mature, experienced teacher; attractive, likeable, and intelligent.

Results of the study, in terms of student achievement, were presented in two forms: informal measures and scores on standardized achievement tests. Some of the informally measured results were (a) active work in the classroom and keen interest in local, state, and

---

[2] Grayson N. Kefauver, Victor H. Noll and C. Elwood Drake. *The Horizontal Organization of Secondary Education*. U.S. Office of Education Bulletin No. 17, Monograph No. 2, 1932, p. 138.

national affairs; (b) an increase of 85 per cent* in the number of periodicals checked from the school library by students in this class over the number checked out to the same group in the previous semester and a much higher group average in reading of periodicals than any other class in school; (c) much voluntary reporting in class of radio programs on governmental matters; (d) active interest in the control of local public utilities; (e) voluntary comments on news from the local papers; (f) 121 per cent more historical books and biographies read by members of this class-group than by pupils of English classes who were not enrolled in this government course; (g) an increased number of editorials and stories on governmental affairs in the monthly issues of the high-school newspaper, all except one contributed by members of this group; (h) inclusion, among points of interest which members of this class wrote the Chamber of Commerce that they wished especially to visit during a school trip to a city some distance away, of the mint, the federal building, the civic center, the state building, and the city police court.

Two standardized achievement tests were administered during the next-to-last week of the semester: the *Nation-wide Every-pupil Scholarship Contest Examination* sponsored by the Kansas State Teachers College, Emporia, and the *Brown-Woody Civics Test.* Norms for the Emporia test were based on a total of 3,277 papers of students who took the examination and, for the Brown-Woody, on 3,508 papers.

Since the Emporia examination was used in connection with a nationwide scholarship contest, it may have been that many of the students who took it were selected from their schools as among those most likely to make high scores and were coached for such an examination. Hence, an entire class taking the examination would normally be expected to make a median score somewhat below the standard median for the examination. How much difference this might make we have no way of knowing, so it is not taken into account in any of the following reports on results of Emporia scholarship-contest examinations. The reader, however, may well keep this point in mind.

The results for this class on the Emporia Scholarship Examination were as follows: of Mrs. Droll's group, two scores were in the high 1 per cent, five more were in the high 5 per cent, two more were in the high 10 per cent, two more were in the high 15 per cent, and the

*Although recent practice seems to be to use "per cent" as one word, for various reasons it was considered desirable to use the traditional form of two words.—Ed.

entire class was in the high 35 per cent of all the more than three thousand students tested. The median for this class was within the high 15 per cent of all students tested. Moreover, there was no special preparation for this examination.

On the *Brown-Woody Civics Test,* of the 3,508 test results tabulated, all but one of the members of Mrs. Droll's class had scores within the upper 35 per cent and the low score was at the 49th percentile. The median for this class was within the high 10 per cent.

It can perhaps reasonably be said that the Emporia Scholarship Examination tended to be mostly factual in nature whereas the *Brown-Woody Civics Test* tended to examine more for understanding of the vocabulary of civics, for recall of the facts of civics, and for application of principles to practical situations.

It is evident, therefore, that this reflectively taught class seemed not only to achieve much of the outcomes which the class was primarily designed to achieve (as shown by the informal measures reported) but also achieved much better than normal expectation in terms of their ability to recall items of factual information and in their retention of understandings. This is the outcome anticipated by those who have carefully studied the nature of reflective teaching and it seems to have been clearly demonstrated by this experiment.

It may be well at this point to anticipate a criticism which is likely to be raised with reference to this and the following studies. It may be objected that no control groups were used which would furnish a basis for comparison and thereby for judging experimental results. It is our considered opinion that such a criticism is ill-founded and will not stand careful scrutiny. In fact, the employed plan was adopted because of recognized faults in what was conventional in educational experimentation before and during the early forties.

The major control in each of these investigations was the group which furnished the standardizing base for each of the tests used. If the standardizing base was reported as representative of the nation as a whole, it was assumed that normal expectation for a normal class conventionally taught would be the standard norm for the test; sometimes the median, and sometimes percentile ranks. Class median scores on intelligence examinations were compared with national medians for the grade-levels in question, such as eleventh or twelfth, in order to reach a conclusion as to whether a given class was approximately above, equal to, or below the standard group. Thus, a group with IQ's approximately equal to those of the standardizing group would nor-

mally be expected to make scores approximately equal to standard on the achievement tests, and variations above or below would be interpreted accordingly.

This plan may not be as neat numerically as certain other plans in use around 1940, but it was thought to be more valid. If, for example, a given teacher teaches both experimental and control classes simultaneously, he or she is bound to be partial to one method or the other. And such partiality is extremely likely to influence classroom results in favor of the favored method, regardless of its effectiveness otherwise. If, on the other hand, different teachers taught the different classes, then the differences in teacher capacity would be likely to outweigh the differences due solely to method. Whatever may have been our progress in methodology up to now, it seems that we still must face the brute fact that a good teacher may out-teach a poor one regardless of the particular "method" he may use. This, however, does not deny that a good teacher with a superior method is practically sure to outstrip a teacher with equally good natural endowments but who employs an inferior method.

It may be suggested that recent developments in the statistical treatment of data might have made possible a better plan than the one used by these investigators. This we neither deny nor admit, for to do either seems unnecessary. In the first place, they were not readily available at the time all but one of these investigations were conducted and, second, the obtained differences were sufficiently large to make minutely refined measures unnecessary. A butcher does not need an extremely accurate chemical balance to weigh out a pound of beef steak.

Three Studies in American History

During 1941 and 1942, three experiments in the reflective teaching of a first semester of American history were reported.[3,4,5] The three teachers were experienced, as with Mrs. Droll were perhaps in their

---

[3] Torry C. Avery. "An Experiment with Teaching a Class in American History According to the Pragmatic Theory." Unpublished Master's thesis, University of Kansas, 1941.

[4] Laura C. Johnston. "An Experiment with the Problem-Raising and Problem-Solving Method of Teaching American History." Unpublished Master's thesis, University of Kansas, 1941.

[5] Roy L. Sailer. "A Third Experiment in Teaching a Semester of American History in Accordance with the Roberts Plan." Unpublished Master's thesis, University of Kansas, 1942.

thirties, and also as with her had taught conventionally prior to the experimental semester which they report. All were able teachers and likeable persons. The Avery study was with a class in a small high school whereas the Johnston and Sailer studies were conducted in somewhat larger, though still small-town, high schools.

Prior to these studies, Lawson M. Roberts[6] had formulated plans for ten units in American history extending from discovery and colonization through the Civil War. The unit titles were as follows:

I. Exploration and the discovery of America
II. Colonization of America
III. The American Revolution
IV. Establishing a national government
V. Winning respect at home and abroad
VI. The Jacksonian period
VII. The Westward Movement
VIII. American society at mid-century
IX. Slavery and sectionalism
X. War for southern independence

All three reports, especially the one by Miss Johnston, describe in considerable detail the way the units actually developed in class. They also included suggestions as to how the various units might be improved. To give even a sketch of those class discussions is out of the question here. We can only say that it was almost amazing to see how closely Mr. Roberts' plan actually worked out in class, even though there was no attempt toward strict following of those plans. In reflective teaching, it is necessary to follow a thought-line as it actually develops in a given classroom. The *study itself* needs to determine the line of study, and circumstances alter cases. Nevertheless, imaginative planning ahead of time can be considered a necessity in which the planner persistently tries to think as students would think under the anticipated circumstances; and Mr. Roberts' anticipations turned out to be highly accurate and usable. Students took hold from the very beginning and participated with interest, even enthusiasm.

The Avery and Johnston classes were small, ten or eleven students each. The Sailer class had about twenty students.

Since each investigator reported findings in a way somewhat different from the others, it will be necessary to take the reports one at a

[6] Lawson M. Roberts. "Ten Units in American History Formulated in Accordance with the Pragmatic Theory of Teaching." Unpublished Master's thesis, University of Kansas, 1941.

time. All three, however, in making comparisons with their own previous experiences with similar classes but taught in a more or less conventional manner, noted distinctly heightened interest, greatly increased voluntary use of library facilities and other reference sources outside of the textbook, increased interest in current national affairs, increasing participation especially by the more reticent students in classroom discussions, increasing tendencies to question statements made by fellow students or teacher or authors, and other characteristics which reflective study is thought likely to promote.

The Avery Study

There were ten members of the class taught by Mr. Avery. His measure of mental ability was the *Army Group Examination Alpha,* given at the beginning of the experimental semester. Intelligence Quotients ranged from 110 to 74, with a median of 92. Since a report of the U.S. Office of Education indicated a median intelligence quotient of high-school juniors and seniors of 108,[7] it appears that this class was of considerably less than average capacity. Whether this measurement was accurate is a distinct question in the mind of the present writer, but it seems probable that the class was below average even though possibly not as much as a median IQ of 92 would indicate.

Two tests were employed to measure achievement. The *Kansas American History Test,* published in 1938 by the Bureau of Educational Measurements, Kansas State Teachers College, Emporia, covered the period of American history included in this study. Form A was given at the beginning and Form B at the close. The median class-score at the beginning was equal to the 5th percentile standard test-score. The median class-score at the close of the period of study was equal to standard percentile 62. Form A was given on September 4 and Form B on the following January 21. On January 24, the *Emporia Every-Pupil Scholarship Contest Examination* was taken and the median score of the class was equal to the standard 75th percentile. Since both of these examinations represented largely the recall of facts in American history from Colonization to the close of the Civil War, the evidence from these measurements is that, in this regard, achievement of this class was markedly above what could reasonably be expected of it. And the class evidently made great advancment during the period of study.

---

[7] Grayson N. Kefauver. *op. cit.*

The Johnston Study

Miss Johnston used more measures of the standardized type than did Mr. Avery. Three tests of mental ability were administered: Army Alpha, Henmon-Nelson, and Terman Group. Army Alpha, given at the beginning of the experiment and representing the entire class of eleven students, gave a median of 103 and a mean of 108. The other two intelligence examinations had been given earlier in connection with the regular all-school testing program and included records of only eight of the eleven class members. Omitted were two who were above the median and one who was below. The averages of those two examinations (Henmon-Nelson and Terman) gave a median of 105 and a mean of 106. Therefore, it seems reasonable to conclude that, in comparison with national standards, the class was about average.

Miss Johnston took 25 weeks, seven weeks more than one semester, to complete the ten units. She had one achievement test which covered the entire time and represented the material covered. This was the *Kansas American History Test* (see above) with Form A given on September 11 and Form B on the following March 4. The median score of the class at the close of the study was equal to standard percentile 50. In other words, on this examination, the class after instruction apparently remembered as many facts in American history as would ordinarily be expected of it.

The *Columbia Research Bureau American History Test,* published in 1926, was also administered at beginning and end of the experiment. This test, however, covered the period from Colonization until 1926. Beside testing the facts of history, this test also seemed to require a considerable amount of reasoning ability and discriminating judgment. However, fifty per cent of the questions related to post-Civil War, a period not covered in the experiment. The median of the class rose from a standard percentile rank of 12 at the beginning to a percentile rank of 37 at the close. When this test was ordered, Miss Johnston did not realize that fifty per cent of it covered material subsequent to what was to be included in the course. But, she administered it anyway and it should be noted that the class gained 22 percentile-points during the study and ended only eleven percentile-points below the standard median, presumably based on study of the entire course.

The *Emporia Every-Pupil Scholarship Examination* was given on January 8, two months before the close of the study. On that examination, the median score of this class was equal to the 72nd percentile of the 176 medians (presumably class medians) reported for the examina-

tion. In other words, the January median for this class was equal to or higher than the medians of 72 per cent of all classes reported. A total of 4,754 scores were reported by the 176 schools used for determining standard test-scores.

On April 8, another such examination was given for which standardization included 4,457 scores from 148 different classes. On this examination the median for this class was equal to the 75th percentile of all medians. Thus, at both times Miss Johnston's class, having average capacity, ranked well above average in performance on examinations requiring primarily recall of factual information.

However, during the year prior to this experiment, Miss Johnston taught an American history class of eight students having a median IQ approximately five points higher than the experimental class. On the January examination the previous-year class achieved a median equal to the 75th percentile of all medians. Of the preceding class Miss Johnston says, "The method of teaching last year was based principally on the Morrison plan which stressed facts and included some drill. There was no drill in the remembering of facts in this experiment." It is evident that Miss Johnston did well, regardless of what method she was attempting to use.

Moreover, during January the experimental group took an annual current-affairs test published by the Searight Publications Bureau of Lincoln, Nebraska. Only two members of the experimental class made scores lower than the score indicated by the publisher as the average to be expected of such students. As Miss Johnston says, "This showing seemed to indicate that our class, although of average capacity, is better informed than average on current affairs."

Among Miss Johnston's informal observations regarding her experience with this experimental class is the following paragraph.[8]

> The problem-solving method of teaching is no panacea; it is no cure-all for the ills of teaching. Members of our group were more willing to hypothesize and, by research, to test out hypotheses and follow through to a conclusion. There was a tendency to do the guessing and let it go at that. This might be expected, since the method was new to the writer and to the pupils. There seemed to be an adequate realization of the problems and perhaps that is all one could expect from beginners. Pupils felt free to express their opinions and to challenge those of others. Discussions were lively and interesting. As was pointed out . . . the pupils raised many questions. For example, "Why didn't we include Lower California in the Mexican Cession?" and "Why did Spain allow only native Spaniards who were Catholics to enter her colonies?"

[8] Johnston. *op. cit.*, pp. 113-114.

> Not all pupils worked up to their several abilities; there were some who gave little and seemingly gained little. However, perfect correlation between ability and achievement may be too much to expect of any method.

Miss Johnston summarized her feelings about reflective teaching as follows:[9]

1. It stimulates wider reading. Pupils read to find the solution to a problem rather than to meet a required assignment.
2. It arouses interest to determine the "why" of things and ideas, and leads the pupils to substantiate their conclusions and opinions.
3. It promotes the recognition of problem situations.
4. It provides opportunity for wholesome "give and take" in the discussion of controversial issues.
5. It seems to result in greater retention of factual information, even without coaching for such retention, than is ordinarily achieved.

The Sailer Study

In the Sailer study, two measuring devices were used in addition to those employed in the Droll, Avery, and Johnston studies: (1) a class group in a nearby community was obtained to serve as a control; (2) a published examination, which represented an attempt to measure reflective thinking and was to a degree standardized, was administered to both experimental and control groups.

As we have indicated in each case, the method used by the three preceding investigators to obtain a "control" to serve as a basis for comparison is, first, to find how the median IQ of the experimental class compares with the national median IQ for students in grades eleven and twelve combined. This, as noted, was reported by the U.S. Office of Education as about 108. Second, the class-achievement medians are compared with nation-wide norms for the examinations by indicating what percentile ranks on the national norm the median score (50th percentile) of the experimental class equals. Thus, when a class of average IQ makes a median achievement-score equal to a national 70th percentile, it is taken to have achieved considerably more than normal expectation.

In addition to such comparisons, Mr. Sailer used a class from a nearby community with which he was familiar and with whose teacher he was acquainted. Of course, Mr. Sailer also reported his own personal impressions, as had the others, and found himself quite in agreement with their observations.

Measurement of intelligence was obtained by use of the *Schrammel-Brannen Revision of Army Alpha, Form A.* The medians of both

[9] *Ibid.*, p. 115.

control and experimental groups were the same, 112, four points higher than what was taken to be the national median for grades eleven and twelve combined. Moreover, the 75th percentiles were nearly the same (122 experimental, 119 control). However, there was considerable difference in the 25th percentiles to the disadvantage of the experimental group (89 experimental, 102 control), there were nine more students in the control than in the experimental group (29 control, 20 experimental). Thus, the Sailer class was about the size of the Droll class and almost twice the size of the Avery and Johnston classes. Moreover, due to difference in policy between the two schools on the age required for entering first grade, students in the experimental group averaged six months younger chronologically than those in the control group. Thus, it seems that, except for the control being a somewhat larger group, the two groups were approximately equal in capacity.

To test for factual knowledge, the *Emporia Every-Pupil Scholarship Contest Examination* was again used, and on January 28 was given to both groups. The standard test-percentiles were determined on the basis of 3,987 individual scores reported from 633 different schools in 29 different states. For the experimental class, the national percentile-ratings of the individual students ranged from 2 to 99 with a 25th percentile of 15, a median of 55, and a 75th percentile of 92. For the control class, individual student-scores ranged in national percentile ranks from 2 to 85 with a 25th percentile of 16, a median of 38, and a 75th percentile of 50. Thus, on fact recall the experimental-class' 75th percentile was seven national percentile points higher than the top score of the control class, the experimental median was five percentile-points higher than the control 75th percentile, and the two 25th percentile scores were about the same, even though the 25th percentile IQ of the experimental group was thirteen points lower than that of the control group.

Comparing, as did Miss Johnston, the experimental-class median score with the class-median scores for the 144 schools reporting, the experimental-class median reached the 67th percentile for all class medians. The median score for the control class was equal to the 30th percentile of all class medians.

These results indicate that, on fact recall, Sailer's and Johnston's results are approximately the same—distinctly better than normal expectation even though no attention had been paid in either class to whether facts were remembered or not. Facts were obtained because

they were needed in problem solving and were so used, but that was all. If Mr. Avery's class was as low in capacity as his obtained IQ's indicated, his class made a remarkable showing.

Mr. Sailer also gave the *Every-Pupil Scholarship Contest Examination* on April 8, some time after the close of the formal experiment. The result for his experimental class was the same as to median (55 in both cases) as the January 28 examination but with a somewhat lower 75th percentile and a somewhat higher 25th percentile. Mr. Sailer suggests the possibility that the drop in the 75th percentile might have been caused by numerous absences from his class by top-ranking students who were drilling for the contest in the work of other classes. His own class was not drilled for the contest.

In order to obtain a standardized measure of reflective ability, Sailer found one instrument which seemed to be at least partially satisfactory—the *Cooperative Test of Social-Studies Abilities,* Experimental Form 1936, by J. Wayne Wrightstone. The test was divided into four parts with available norms based on the scores of 1,250 students. Part I dealt with obtaining facts, Part II with organizing facts, Part III with interpreting facts, and Part IV with applying facts. The total score was found by adding together the four part-scores. The same form of the test was given twice, once at the beginning of the school year when the experiment was started and once near the close of the school year, some time after the close of the experiment. It was given to both experimental and control groups at about the same time. It was expected that a different form would be given the second time, but, since no second form was available, the first had to be repeated. However, during the entire intervening period nothing whatsoever was said in either class about the test or that it would be given again, so that the practice effect in favor of the second would probably not have been very large, if any. We merely mention this test here to show that an attempt at such type of measurement was made. The very make-up of the test was such as to be rather unsatisfactory both to the investigator himself and to his advisor, the present writer. Those four factors may indeed be aspects or elements of reflective thought, but it seems hardly sound psychologically to assume that a true measure of reflective ability can be obtained by a summation of scores which purported to measure those four factors in separation from one another.

For what they are worth, however, the results of this test are given. At the beginning of the year, for some reason, the control group scored considerably higher than the experimental: for the control group, the 75th percentile was 155, the median 115, and the 25th percentile 90. For

the experimental group, the 75th percentile was 156, the median 89, and the 25th percentile 60. Near the close of the year, for the control group the 75th percentile was the same, 155, the median was seven points higher, 122, and the 25th percentile was four points lower, 86. For the experimental group near the close of the year, the 75th percentile was 19 points higher, 175, the median was 52 points higher, 141, and the 25th percentile was 20 points higher, 80. Thus, according to this measuring instrument, during the year the control group made little progress in reflective capacity while the experimental group was improving considerably. The announced norms for the test, based on 1,250 cases and for the *close of the senior year,* were 75th percentile 162, median 141, and 25th percentile 119. Since it is not known what type of students composed the 1,250 to whom Dr. Wrightstone administered his examination for standardization, we have no way of knowing or of surmising why the results for both classes participating in the study were no higher comparatively than they were.

Two Studies in Upper Elementary Grades

The Trefz Study

The foregoing studies dealt with classes in senior-high school. The two remaining deal with upper elementary, the Trefz Study[10] with Grade VI and the Reader Study[11] with Grade V. Moreover, because in these grades a single teacher was in full-time charge of each class, the experiments cover all of the subjects studied by these pupils and for almost an entire school year at a time. And, in addition, two successive years of work are included in each report which comprised the master's thesis. Furthermore, the Trefz study can here be extended for four additional years (six years in all) due to informal reports made to the present writer by Mrs. Trefz at the close of each year. Thus, evidence is given as to whether a teacher may profit by experience in reflective teaching and secure better results a second, third, fourth, fifth, or sixth year than were secured on a first attempt. Both teachers had not previously attempted reflective teaching as such, and have more than once indicated that they had been largely unaware of it,

[10] Ida R. Trefz. "An Experimental Study of Teaching a Sixth-Grade Class for Seven Months Employing a Pragmatic Program." Unpublished Master's thesis, University of Kansas, 1941.

[11] Edna C. L. Reader. "An Experimental Study of Reflective Teaching in a Fifth-Grade Classroom for a Two-Year Period." Unpublished Master's thesis, University of Kansas, 1953.

other perhaps than the ordinary conviction that we ought to do what we can in helping children learn how to think.

In order to give the reader an idea of what led to Mrs. Trefz' study, we shall let her tell it.[12]

> The method used by the writer in teaching previous to this experiment was confined more to factual learning, the text book method being used. Questions were asked and answers were given, with little discussion. Reports were made by class members and accepted by the others without agreement or disagreement. Maps were used for memorizing locations, not as sources of information for continual reference in solving problems. Rules were memorized; not formulated or "arrived at" by the pupils themselves. Particularized learnings was the major objective; not development of an enhanced ability to learn for one's self. School and life were not as closely related as they might have been, even though field trips were taken and community projects talked about.
>
> Then, during the summer of 1939, through the study of teaching procedures and through reading books by Dewey, Bode, Kilpatrick, Wheeler, and others, ideas and words such as democracy, discussion, agreement, harmonization, enhancement, relationships, problem-solving, and reflective thinking began to take on new and different meanings. In the fall of 1939 the pragmatic program became the basis for our teaching procedures.
>
> I had this same group the year before (1938-39), having been promoted from the fifth to the sixth grade with them. Unconsciously, in September, 1939, my teaching was changed to fit a somewhat higher level of mental capacity, because I changed my objective to reflection-level thinking. Before, it had been recollection and recognition (understanding) levels.

It might also be noted here that, after Mrs. Trefz had taught this class in fifth grade, it came with her to sixth grade somewhat below the achievement-norm for beginning sixth grade. As a result, however, of her teaching this same class during sixth grade but reflectively rather than conventionally, the class finished the year above normal in achievement. This is a rather significant comparison because the only factor which differed materially was the method of teaching. Both years it was the same class and the same teacher. Moreover, when the teacher taught conventionally, she was convinced that that was the way to teach and had been teaching that way for a number of years. The altered teaching method came *after altered convictions* regarding teaching. Hence, was avoided the difficulty, when a teacher is simultaneously handling two different classes in two different ways, of the teacher's conviction that one way is better than the other and of consequent favoring of the preferred method by the very fact of personal conviction. It is for this very reason that the present writer has advised

---

[12] Ida R. Trefz. *op. cit.*, pp. 3-5.

experimenting teachers not to use classes taught by themselves as control groups.

Mrs. Trefz tells as follows of her first experimental year.[13]

Group intelligence and achievement tests were given. In the 1939-40 class of thirty-six sixth graders, the I.Q.'s ranged from 69 to 143. The median was 97. There were twelve pupils whose I.Q. fell in the normal range (90-110), eleven were above normal (111 up), and thirteen were below normal (up to 89).

Results on the achievement tests taken in October seemed to indicate that teaching was still pitched at too low a level to challenge the better students because the thirteen pupils ranking below normal in intelligence were achieving more in proportion to their ability than those with the higher I.Q.'s. This conclusion was reached after reading an article by Bayles.[14] He says, "We believe that the interests of a class as a whole are best served when teaching is pitched above the median of the class, say at the level of the second quartile, and made sufficiently flexible to challenge the best pupils. If the more competent members of a class, usually the most influential in determining the general class attitude, are convinced that a class is worth while, the remainder usually follow along and gain considerable benefit from the instruction, whether they comprehend all that goes on or not. The writer's general impression of teaching today is that it is focused upon too low a level to present a wholesome challenge to pupils."

Hereafter teaching was pitched on a higher level. Class work was made more challenging for the brighter pupils, yet interesting to the duller ones.

In the spring of 1940, Form B of the achievement test, of which Form A had been used in the fall, indicated that those whose I.Q.'s lay in the range of normalcy (90-110) for their grade made an average of nine months advancement in the five months between tests, those whose I.Q.'s lay in the above normal group (111 up) made an average of seven months and a half advancement in the five, and those in the below normal group (up to 89) made an average of ten months advancement in the five.

Should the teaching have been pitched at a still higher level, or was the test a satisfactory criterion for judging the advancement of the higher group? We were bothered by this question. The test covered practically the same material as Form A on which these higher pupils had made good scores at the beginning of the study. In order to have secured a complete picture of the gain of all groups, we felt that another test should have been given the superior group to test the depth of their gain.

The children of this class (1939-40) did come to know democracy as more than a form of government. They enjoyed attacking problems, searching for data, discarding irrelevant material, and then accepting solutions that were most in keeping with available data. Sensing that a degree of success had come with this experiment, yet feeling it an unfinished task, we decided to continue using the pragmatic program into 1940-41, profiting from the experience gained in 1939-40.

---

[13] *Ibid.*, pp. 5-8.

[14] Ernest E. Bayles. "The Relativity Principle as Applied to Teaching." *University of Kansas Bulletin of Education*, 4 (Feb., 1940) 10.

The present writer can well remember Mrs. Trefz' concern with the relatively poor showing of her upper-IQ group in terms of progress during the period between tests. To have the high-capacity group registering only 7½ months improvement in five while the low-capacity group registered ten months improvement was to her a shock. For, if IQ means capacity for rate of growth, then with equal treatment or opportunity a high-capacity group will surely register more advancement than will one with low capacity. When the opposite occurs, something must be holding back the high-capacity group. Mrs. Trefz finally decided that her teaching was still not pitched on a level high enough to give the upper two-thirds of the class the challenge it needed. Hence, during the second year, she sought to have her class delve more deeply and study more difficult problems than they did the preceding year. Space cannot be taken here to give descriptions of the studies as they were pursued by Mrs. Trefz' class in the second year of her experiment. She gives such descriptions in her thesis.

In presenting the test results for the second year of Mrs. Trefz' experiment, though not quoting exactly we shall follow very closely her own report. During the first year, 1939-1940, Forms A and B of the *Metropolitan Achievement Tests, Intermediate Battery, Complete (Revised),* were used: Form A in the fall and Form B in the spring. In the second year, 1940-1941, Forms C and D were used, Form C in October and Form D the following April. The tests were given approximately seven calendar months apart. Intelligence quotients for the class, taken from school records, showed a high of 127, a low of 63, and a median of 105. Thirty pupils composed the class. As shown by the October achievement test, the class was then at grade-level—six years and one month (6.1)—as shown by the calibration chart for the test.

In April, almost seven months later, the class median was found to be *five months above* grade-level. Therefore, since advancement of the class was five months more than the seven which was normal, it is evident that median class-achievement showed twelve months improvement in seven.

The largest gain made by any class member during the seven-months period was 15 months and by the student having the highest IQ, 127. Gains of 14 months were made by four, having IQ's of 121, 108, 87, and 64, respectively, the latter being a repeater in the grade. Gains of 13 months were made by two, with IQ's of 105 and 110. Twelve-months gain each were made by eleven students with IQ's ranging from 121 down to 85. Those having IQ's from 111 up made a

mean gain of 11.6 months during the seven, those with IQ's from 90 to 110 made a mean gain of twelve months, and those with an IQ of 89 and below a mean gain of 10.3 months in seven. The mean gain of the entire class taken together was 11.3 months during the seven months between tests.

It would seem, therefore, that pitching the level of difficulty higher the second year than the first did accomplish the anticipated result—the middle group outgained the low group by almost two months whereas the preceding year the low group had gained ten months while the middle group was making a gain of nine. Moreover, the upper group this year outgained the low group, although by only 1.3 months on the average whereas during the first year its gain was only seven months while the middle group was gaining nine and the low group ten. However, the showing by the upper group was still unsatisfactory since its mean gain during the seven-months period was .4 month less than that of the middle group although it should seemingly have been perhaps at least two months more than that of the middle group.

Nevertheless, the over-all showing of the class was indeed remarkable. This was a widely used test, well standardized, and calibrated in terms of ten-month years. That is to say, the median showing of the class—twelve months improvement in seven—means that the class on the average improved 1.2 years during .7 year. In other words, the median gain of the class was about 72 per cent more than normal expectation. And it is very clear that the over-all gain of the class was not at the expense of the low-capacity members, for their mean gain was 47 per cent more than normal expectation for *average*-capacity pupils.

It would seem that this showing by low-capacity class members deserves emphasis, due to common conviction and indeed frequent claim that reflective teaching (or teaching so as to make pupils think) may be fine for the best students in a class but it will work a hardship on average and poor students. Mrs. Trefz' experiments for both years clearly showed almost the opposite, that lower-capacity students were profiting from reflective teaching more than were higher-capacity students.

Thus, at the close of two years' experience with reflective teaching, this teacher was still wrestling with the question, "What must I do to enable my best students to achieve as nearly up to their capacities as my poorer students do?" In studying this problem, Mrs. Trefz analyzed improvements made on the various parts of the test—that is, in the various subjects which were covered—and found that at the begin-

ning of the year the 1940-1941 class was below grade in both arithmetic and spelling. Regarding this analysis she had the following to say:[15]

> Because of the very low median on the fall tests in arithmetic and spelling, special effort was made to help the children do reflective thinking in these subjects. Many questions were asked both by the teacher and by the pupils. Searching of data for the best answers kept the children interested, alert, and wanting to learn. Why they were missing problems, and not spelling words correctly, became individual problems. Each pupil was encouraged to find his trouble. Often the class helped him, and then, sometimes, the pupil had to have the teacher's help to straighten him out. This was often done in a supervised study period, the period being provided, not at fixed times, but when a "felt need" seemed to indicate it. The children enjoyed the supervised work period, and often asked questions on what had been bothering them. Many times during the class recitation periods the children recognized their need for more problems in the different fundamentals and asked for drill in these.
>
> In spelling, visualizing the word, pronouncing it correctly, being aware of the syllables, noticing silent letters, and spelling orally, helped the children feel an answer to the question: "How does the word *have* to be spelled?" They enjoyed a baseball game of spelling very much, the game being played like regular baseball, using words for balls.
>
> Indeed, it was gratifying to discover the results as indicated by the spring achievement test in these two subjects especially. The median for the class in average arithmetic went from seven months below grade to four months above grade, indicating, in all, eighteen months gain in seven months time—the seven months below grade, the seven months interim (between tests) and the four months above grade. In spelling there was a like result, from six months below grade to three months above grade, sixteen months gain in all.

Mrs. Trefz' analysis led her also to make the following comments regarding the social-studies aspect of the year's showing.[16]

> In the sixth grade we have social studies instead of history and geography, and even though the class was most interesting, and social studies perhaps the best adapted of all the subjects for reflective thinking, the tests in geography and history did not cover social studies subject matter. Because of the nature of the tests, being of the multiple choice type of objective tests, they required mostly recognition-level thinking (understanding) and not reflection-level thinking. As a result, we believe that the outcomes of some of our very best work are not reflected in our test results at all. We believe that if our test battery had included one in the social studies proper, the class would have shown high level achievement in that test.
>
> Throughout the year there have been unmistakable indications of an increased ability to think straight. The members of the class see their prob-

---

[15] Ida R. Trefz. *op. cit.*, pp. 61-62.
[16] *Ibid.*, pp. 62-63.

lems more quickly, and are better able to use the facts they possess in arriving at correct generalizations.

Following are the conclusions of this experimenter at the close of two years' experience with reflective teaching.[17]

> We have conducted classes the last nine months, using to the best of our knowledge pragmatic teaching, which seeks to promote reflection-level studies of ideas in contrast or conflict, looking toward development of enhanced and more harmonious outlooks on life.
>
> 1. We found it to be very interesting both to pupils and teacher. The children enjoyed the privilege of discussion. The teacher needed to keep up with the affairs of the day, to be able to contribute her share in discussions.
>
> 2. We found changed attitudes. The class became more cooperative. The children were led into respecting the rights of others, not only in the room, but also on the playground and at home.
>
> 3. We found enhanced and more harmonious outlooks on life. The rendition of the play "I'm an American" revealed this enhanced outlook. Through the study of the European countries and contrasting them with America, the children realized what it meant to be an American. Because of this realization very little coaching was necessary in putting on this play successfully. The children's art project taught them that things are neither good nor bad in and of themselves, but in the relationships in which they are seen. The study of budgeting both time and money was shown as carrying over into the home by reports of parents who remarked about instances of planning and budgeting at home.
>
> 4. We found improvement in self-reliant, reflective thinking. Radio news of the day were sometimes discarded as propaganda and reasons given why they were evaluating them as they listened, and the next day shared what they thought was worth while.
>
> 5. We found marked improvement in subject matter knowledge. In 7 months time the class made 18 months progress in arithmetic, 16 months progress in spelling and an average of 11.3 months progress in all subjects.
>
> 6. We found that pitching teaching on an upper level of class ability increased the achievement of all groups; low as well as middle and high capacity. In 7 months, the ten pupils with the highest I.Q.'s in the class (112-127) made an average of 11.6 months improvement, and the ten pupils with the lowest I.Q.'s (63-100) made an average of 10 months improvement. Moreover, the higher the level of ability aimed at in the teaching, the higher the correlation between ability and achievement on the part of all groups; high, middle, and low.

Mrs. Trefz' formal report, which represented her master's thesis, was made in 1941 after two years of experience in reflective teaching. But her interest was aroused and she continued her measurement program for four more years, each year using one form of a test battery in the fall and a different form in the spring, six or seven calendar

[17] *Ibid.*, pp. 66-68.

months later. In some years the Stanford test was used and in others the Metropolitan. Each year she would send to the present writer an informal, though careful and accurate, report of her findings. These reports are on file in his office.

In her *third year,* 1941-1942, the tests were given in October and March, six months apart. The median grade-placement in October was 5.8, or the eighth month of the fifth grade, which was at that time three months below the grade-standard of 6.1. The results of the March test showed median grade-placement of 7.8, or eleven months above the standard grade-placement of 6.7. This year her class was made up of 25 pupils, and its gain was more than 15 months in median grade-placement during a six-months teaching period.

Average (presumably mean) gains in grade-placement made by each IQ group for the six-months period between tests were as follows:

Above normal IQ (111 up)	17 months
Normal IQ (90-110)	18 months
Below normal IQ (up to 89)	9 months

Thus we note that, in her third year of reflective teaching, this teacher—a mature and experienced teacher at the outset—made a large gain over her second year, an average gain of three times normal expectation or eighteen months' improvement in six. During this third year her below-normal group continued to do approximately as well as the first two years but her middle and upper groups did much better than before. Yet her top-capacity group still fell short of her middle-capacity group regarding gain during the period.

During the fourth and fifth years, results continued on this remarkably high level, even showing some gains. Mrs. Trefz' final report to the writer was for the year 1945-1946, her sixth year of reflective teaching. In this case the tests were given about six months apart and the gains were as follows:

Above normal IQ, average gain	21 months
Normal IQ (90-110), average gain	19 months
Below normal IQ, average gain	14 months

This was with a class of 25 pupils. The class started with a median of two months below standard grade-placement and ended with a median of eleven months above standard grade-placement. It is evident that this year the objective of getting students to achieve approximately in accordance with their several capacities was accomplished. The average of the top-capacity group was two months higher than

that of the mid-capacity group and its average in turn was five months above that of the low-capacity group.

Thus, it is seen that a teacher, even though mature and experienced at the outset, can, by getting hold of a fruitful idea or plan, greatly improve the achievements of her pupils as measured by available standardized tests, and can continue steadily to improve those results over a period of at least six years. This is indeed a splendid record of achievement and it is evident that this teacher is deserving of highest commendation for her efforts. The question, of course, arises as to whether others could do the same. The findings of the Droll, Avery, Johnston, and Sailer experiments appear to give an affirmative answer, although for only one year. In addition, we have one more experiment to report, this one terminating almost three years prior to publication of the present paper.

The Reader Study

Miss Reader is a mature and personable young woman, younger than Mrs. Trefz, teaching fifth grade in a mid-western town in the 5,000-15,000 population group. This was a two-year study, performed during 1951-1952 and 1952-1953.

As with Mrs. Trefz, Miss Reader had to follow essentially the course of study required in her school system. Modifications could be effected only within that general framework. As she says, "The units of reflective thinking have been worked out following the interests of the pupils, guided by the teacher so that required material in the course of study would be covered."[18] Later she enlarges,[19]

> The course of study follows fairly closely the basic textbooks which are provided for each child. There was a definite shortage of supplementary and reference material, but some was provided by the school during the year and the County Librarian was generous with library books, checking out twenty-five or thirty at a time for use over two-weeks periods or longer if needed.
>
> The children were accustomed to the traditional type of teaching and had learned, as most children do, to take textbook and teacher's word as final. An effort was made to lead them into thinking their own way through the questions they normally raise, and to bring to light incompatibilities in their outlooks. It was rather hard to do this, for many children seldom expressed their own ideas in the classroom for fear of being laughed at or thought impertinent, or just from force of habit. But, after the teacher had said several times, I'm not sure, let's find out, they began to feel that perhaps after all their ideas were worth expressing.

---

[18] Edna C. L. Reader. *op. cit.*, p. 7.
[19] *Ibid.*, pp. 32-33.

The writer had, previous to this study, been using broad social units for two or more years so decided to continue this method, adding to it reflective teaching. At first the children did not notice much change and a casual observer who might have stepped into the classroom would not have been likely to notice much difference. The subject matter was related directly to everyday experiences outside of the classroom so that it would have meaning to the pupils. Then, as they followed the trend of thought from the familiar to the unfamiliar, questions were asked about things outside the realm of their knowledge, and incompatibilities in outlook were brought out. At first this only astonished the pupils for, when they could not find an immediate answer, they turned to the teacher and said, "What *is* the answer?" When they were not told but were challenged to find their own answers, their next question was, What shall we do now?

But they were instantly alert and eager to meet the situation. Even the slower pupils, who were accustomed to sitting quietly with their minds on something else until the correct answer was given by better students, looked up in surprise at this turn of event. If the challenge or conflict had been planned ahead of time by the teacher—and most of the first ones were—material was made available in the classroom so that each child could find information on the subject under discussion, at his own reading level.

Earlier Miss Reader says,[20]

Of course, in most instances the teacher knows what the answer will be, since she has seen the data tested many times before. But the children should be permitted, within suitable or reasonable limits, to find the answers for themselves. Thus they will learn not to accept just any theory that is advanced, but will insist that it be tested. And they will gain a feeling of confidence in their ability to find answers for themselves.

Miss Reader has this further to say,[21]

The incompatibilities in the pupils' outlooks which were brought out in the daily lessons led far from the textbooks, although such were still used for reference work. They were, in fact, the best reference materials available and it was easy for the teacher and pupils alike to slip back into the habit of accepting statements in these books as correct and final, without challenge. But, by constant questioning, the habit was formed of checking to see whether the children had beliefs or opinions differing from those stated in the text, then checking other material for additional data. The children also learned to form hypotheses or, as they said, to guess at the answer, then defend such hypotheses or guesses until proof was found that they were not tenable. Of course, at first the guesses were inaccurate and many times unreasonable. But, as the habit of testing these opinions grew, the youngsters found that, if in the first place they considered the problem from all the angles which they could see, their guesses were more nearly correct.

Reflective thinking *had* to be used when the data on hand were under consideration. Memory alone could not help a pupil decide which of the

---

[20] *Ibid.*, pp. 13-14.
[21] *Ibid.*, pp. 35-36.

hypotheses he had formed was most in harmony with the data. Not only did the children face problems, they also gained insights into how to work out solutions for themselves.

The 1951-1952 class had an enrollment of 34 pupils, an IQ range of from 75 to 131, and a median IQ of 102. The mean IQ was 98. The 1952-1953 class was composed of 34 pupils with an IQ range of from 81 to 141 and a median of 106. The mean also was 106. Both years the achievement test used was *Metropolitan Achievement Test, Intermediate Battery,* with Form S given in October each year and Form R given the following April. IQ's were taken from the permanent record cards of the students and were determined in the course of the regular testing program of the school.

For the first year of the experiment (1951-1952), the October achievement test showed the class to be a little below standard grade-placement—4.8 instead of 5.1. Both the median and the mean for the class was 4.8.

When the second form of the test was given seven months later, the mean grade-placement was 6.3 as compared with the standard at that time of 5.8, or five months above grade-standard. The median again corresponded to the mean. Thus, in the seven-months period between tests, the pupils had gained 15 months in grade-placement as compared with standard expectation of seven months. Since, as explained before, grade-placement on this test is computed on the basis of a ten-months year, this gain is a bit more than twice normal expectancy.

During this year the high-capacity group (IQ's above 110) gained an average of nineteen months during the seven months between tests; the mid-capacity group (IQ's between 90 and 110) made an average improvement of twelve months in seven.

It is evident that, even during her first year of reflective teaching, Miss Reader's class not only greatly exceeded normal expectation but also made gains which correlated highly with pupil capacities, the high group averaging four months more than the middle, and the middle averaging three months more than the low. Miss Reader had read of Mrs. Trefz' experiences.

The 1952-1953 class was evidently an unusually capable one. In October when standard grade-placement was 5.1, this class made a mean grade-placement of 5.5 and a median of 5.8. The median grade-placement was ten months, or one year, higher than the median grade-placement of the previous class. Only in arithmetic and spelling was

this class below grade-placement during the second month of the experimental year.

When the second form of the test was given seven months later, late in April, the class achieved a mean grade-placement of 7.6 and a median of 7.9. At that time standard grade-placement was, of course, 5.8. Thus, using either mean or median grade-placement in comparing October and April tests, this class made an average gain of 21 months during the seven-months period between tests. This is a gain of *three times normal expectancy*. In this case, the high-capacity group made an average gain of 23 months in seven, the mid-capacity group an average of 19 months in seven, and the low-capacity group an average of *17 months improvement in seven.*

Test results for achievement gains shown by pupils for the two years of the Reader experiment are summarized as follows:

	In 1951-1952	*In 1952-1953*
During seven months, the		
mean class gain was	15 months	21 months
median class gain was	15 months	21 months
high-capacity group mean gain was	19 months	23 months
middle-capacity group mean gain was	15 months	19 months
low-capacity group mean gain was	12 months	17 months

The *Metropolitan Achievement Test, Intermediate Battery,* was used both years. Form S was given early in October and Form R late the following April.

I cannot refrain from speaking further of what this second class did. In the first place, it might be thought that a class which started above grade-expectation would hardly be expected to make as much improvement with altered teaching as one which started below grade-expectation. This is a feeling rather commonly held even though careful thought shows it to be theoretically fallacious; with equal treatment a faster group should progressively widen a gap between it and a slower group. This class, although beginning the experimental year a full year in advance of the first class, outstripped the average gains of the first-year class by six months. Looking at it another way, the class, although starting approximately a half-year above grade-placement, made a gain during the seven months between tests of 300 per cent normal expectancy.

It should also be noted, and emphasized, that the mean grade-placement of 7.6 for this class in April was a very conservative rating since many of the individual grade-placements had to be below the

scores actually made by the pupils. The reason for this is that the highest grade-placement given on the conversion scale for the test was 9.0, whereas test scores could be and were made which were higher than a grade-placement of 9.0. Of the 34 test scores made on the science part of the test, 17 could not be charted at their actual value because they had a grade-placement higher than 9.0. In addition, there were 16 scores in geography, 13 in history, ten in literature, eight in English, two in arithmetic-problems, five in vocabulary, and ten in reading which ran off the top of the chart and had to be computed as only 9.0 when they actually were higher—81 in all. How much higher than 7.6 the class mean would have been if the grade-placement chart had been calibrated above 9.0 is a matter for conjecture. However, it should be noted that the final median of 7.9 was doubtless closer to the true central tendency of the class than was the mean of 7.6 because, as statisticians understand, of the tendency of atypical scores at the extremes of the range to influence a median much less than they do a mean.

Miss Reader terminates her report of test results with the following interesting item:[22]

> The children in the class the first year of the experiment, 1951-1952, were given an achievement test at the beginning of the next school year, 1952-1953, when they were in the sixth grade with a different teacher. The Stanford Achievement Test, Intermediate Battery—Complete, Form G, was both given and checked by the elementary supervisor. The subject-matter fields of this test were similar to those of the Metropolitan Achievement Tests which had been used for fifth-grade testing. However, both the tests and the conditions for the test were different from those of the previous year.
>
> The results of the Stanford Achievement test showed that the median grade-placement for the class was 6.1. Since the class median grade-placement of the Metropolitan Achievement test given the preceding April was 6.3, this was a difference of two months on different kinds of tests, one given before and one after the summer vacation. Compared with the median grade-placement at the beginning of the fifth-grade year, the children still had a median gain of thirteen months. And, compared to the average grade-placement of all the schools in the city, they were five months above the system average.
>
> This appears to indicate that knowledge gained by reflective teaching is not easily forgotten and is applicable to different kinds of tests.

It would indeed be interesting, and profitable as well, to know what happens to youngsters after an experience of this kind when they return again to conventional teaching. Regarding this question we have no data, although Mrs. Trefz once reported to the writer that a

[22] *Ibid.*, pp. 99-101.

teacher who had her pupils in a following grade spoke to her of the rather unusual interest and capacities of the pupils. The real question is, of course, not what youngsters would do when returning to conventional teaching, but what might reasonably be expected from pupils who had had teaching of this kind during their entire public-school and even college experience.

In completing our report on Miss Reader's experiment, I should let her speak for herself. Following is her entire chapter entitled, Summary and Conclusions.[23]

> There are two steps in reflective teaching, problem-raising and problem-solving. The important object of the whole program is to challenge the children with an I-don't-know situation so they want to find the answer, then help them realize that they must consider all relevant data they can find before they decide which answer is correct. The children are led to question and compare, instead of just memorize facts which the teacher or the textbook presents to them.
>
> It is not an easy way to teach as far as the expenditure of energy is concerned. It is far less work to make page-by-page assignments for the children to read and to accept without question. But the results of teaching reflectively are so much more satisfying that, because of this satisfaction, it might in the end be considered the easier method. To see children develop the ability to think through a problem, to see their interest and determination and their looks of satisfaction when they have met the challenge, is worth more to a teacher than the extra effort put forth in making and carrying out plans to give them the challenge.
>
> During the two years of this experiment, in a fifth-grade classroom, the results have been most satisfying. In the first place, during a seven-months period between tests, the median gain in achievement—fifteen months the first year, or two times normal expectancy—would be enough to merit continuation of this type of teaching. But even more important is the all-round development of the children. No grade cards are given in this school system, but parent-teacher conferences are held. Many parents voiced the opinion, I don't know just what you are doing, but my child really likes school this year. And he seems to be learning, doesn't he? Usually this question was asked because there is still a tendency on the part of the public to doubt whether children can learn quite as much when other than traditional methods of teaching are used. When given a brief explanation of teaching methods in use and of consideration given to development of the whole child, then given actual proof of the gains their children were making, the parents were more than satisfied.
>
> As far as the children themselves were concerned, they improved in their ability to think clearly, and enjoyed the work. School no longer seemed to be an isolated portion of a day, for home experiences were used at school and school experiences were used at home. The children had

[23] *Ibid.*, pp. 102-105.

come to realize that everyone has a right to his own opinions; that no answer should be scoffed at for it might prove, when tested, to be the correct one. They learned that there is a reason for doing most things as we do them, and that those reasons can be found if we shall but search for them. Some pupils accepted the challenge of reflective teaching better than others, but even those who did not always respond gained in many ways.

The question might be asked, Was drill or practice ever necessary on fundamental combinations in arithmetic? Only after the pupils realized the need for practice was it introduced. They were never given drill just to learn a certain combination or process. But this type of teaching helped a child to realize his weakness. One child aptly remarked, I knew how to work that problem the same as the others did, but I just missed my multiplication. Boy, I'm sure going to learn that tonight. There was no need for the teacher to require drill.

The role of the teacher as a guide was a pleasant one. Sometimes the problem became so large that the children became discouraged rather than challenged. It was then that the teacher needed to help make the problem more simple by advancing theories which would lead toward solution and start the children thinking along profitable lines. She worked only as a group leader, so was never afraid of losing the respect of the children by admitting that she did not know. She was just another person and had a right to make a mistake once in a while.

There were very few disciplinary problems in the room either year, and certainly *no* serious problems. Of course, the children were normally cooperative and, had traditional methods been used, there would probably not have been many problems. However, in several cases it is quite certain that the challenges given the pupils changed their attitudes toward school and perhaps also effected better adjustment in their social life in school.

The interest, sparkle, and enthusiasm which could be seen on the faces of the children as they started each school day and attacked new problems; the acceptance of each child as an important member of the group; the growing confidence in ability to meet new situations successfully both in school and out of school; the improved ability to make right guesses after considering a problem; and the actual gain in achievement in subject-matter fields; all combine to provide evidence that the reflective method of teaching is far superior to traditional methods. Moreover, it would seem to be the way to achieve democracy in the school room.

General Comment

In this survey six experiments in reflective teaching are reported. All of the experiments were conducted in connection with master's theses at the University of Kansas under the advisorship of the present writer. All are hitherto unpublished. The first four dealt with social-studies classes in high school, grades eleven and twelve, the fifth with sixth-grade classes, and the sixth with classes in fifth grade.

The major question raised in each case was how well pupils do in school work when taught reflectively as compared with achievement in classes taught conventionally. Reflective teaching is described and illustrated as that which starts with a problem—an "I-don't-know" situation or one in which the class members are faced with a question which they cannot answer, yet would like to answer. Hence, *problem raising* is the first step.

Problem solving then involves, first, formulation of possible solutions by the students; second, logical deductions as to what practical consequences the various proposed solutions imply; and third, testing the various proposals by determining which, if any, is successful in deducing correctly (a) data which already are known and (b) data which are not already known but have to be discovered by search, or produced by experimentation. Thus, reflective study begins with a question and, after study, ends with an answer to the question. Facts are gathered in profusion but not for the purpose of memorization and later recall; they are to be used only in solving problems. Whether they are remembered is a matter of no immediate concern for a teacher whose major concern is promoting reflective ability. In none of the foregoing experiments did the teacher concern himself or the class with the recall of factual information. Few, if any, fact-recall examinations were given during the experimental teaching.

An experimental question of each and all investigators was, however, "How well do the members of classes taught in this manner compare with those taught conventionally with regard to what is covered in conventional, standardized examinations?" Analysis of such examinations shows that they test predominantly for fact recall, together with some testing for understanding of principles. The results of formal tests which are reported in this paper answer, for the most, these questions only. In other words, they tell whether the members of classes taught reflectively recall facts and retain the ability to use ideas or principles as well as, better than, or not as well as members of classes taught conventionally.

As far as these six experimental studies supply an answer to this question, the answer is unqualifiedly that students taught reflectively do better—a great deal better—than students who are taught conventionally, even when, as is doubtless shown in all cases wherein the Emporia Scholarship Examinations were used, conventionally taught students are given considerable coaching directly focused upon the examinations.

Statistical devices to show the significances of obtained differences have been quite unnecessary in these reports because the sizes of the differences have been so large as to make their significance obvious.

A second point, evidenced mostly by informal, yet careful observations, was that objectives particularly sought in reflective teaching gave reasonably dependable evidence that considerable progress had been made; that students actually made distinct and more than normal progress in achieving the kind of self-reliance in the face of difficulties which reflective teaching would be expected to engender. If it is suggested that these measures were "purely subjective," we would ask a reader to examine the basis for each such conclusion, noting the actual observations that were made and the care which the observer used in stating what conclusions he reached in light of the observations which he made.

A third major point to be noted is the one, indicated by the Trefz and Reader studies (particularly the Trefz study), that even an experienced teacher does not do as well during the first year of reflective teaching as, with reasonable effort, she or he can expect to do in subsequent years. Mrs. Trefz' classes showed steady achievement-gain over a period of six years and Miss Reader's second class made considerably more gain during the experimental period than the first one, even though the second class came to her one full year in grade-placement higher than the first class. It is reasonable, therefore, to expect that the first four teachers reported herein would also have been able to show greater gains by their classes after they had been teaching reflectively for two or three years.

A fourth point to be noted, although not emphasized previously in this paper nor in the reports on which it is based, is that many benefits of reflective teaching may be achieved within the framework of courses and of school organization already in existence. No individual teacher has to upset a given school or its curriculum in any way, yet can greatly improve his own teaching. Doubtless it would help to have other teachers doing the same because of the enhanced effectiveness of cooperative endeavor. But it should be remembered that in each of the foregoing investigations a single teacher was working within the conventional framework of subject-matter and school organization and was working by himself. This fact might be remembered when one reads of the results obtained through many of the curricular innovations which have of late become popular in certain quarters and have been widely heralded as modernization of the curriculum.

Finally, as to whether a teacher can stay in a community after conducting classes in such way as to promote reflective thinking on problems which are challenging to their class members—problems which frequently represent matters distinctly controversial in the community—the answer with reference to Mrs. Trefz and Miss Reader is that they are still teaching (as this is written) where they were during the years reported herein. As to the other four, the writer knows of two only. During the experimental year these two experienced no difficulty in the community where they taught, and later moved by their own choice to better positions. And it is to be expected that genuinely reflective, democratic teaching tends not to stir up a community even though controversial matters are considered because all conclusions and convictions reached by class members are those which the studies indicate are necessary; not those which a teacher brings to a class and proceeds to indoctrinate, whether in light of evidence which is favorable or unfavorable. I once had one of my students—a social-studies teacher, also superintendent, in a very small town in rural western Kansas—say to me, "Teaching reflectively, I don't need to have the slightest hesitation in having students out here study anything that they ought to study; even labor unions." The point is that, in this nation, teachers who understand how to teach democratically and do so in a competent way are likely to have *less* trouble with a community than other teachers and will probably be able to stay in their jobs as long as they wish to do so.

This is not to say, however, that reflective teaching—or any other kind, for that matter—will of itself be guarantee for holding a job. Ill considered and thoughtless actions, exaggerated claims or much discussion of what one is doing or attempting to do, exhibitionism or rambunctiousness, unnecessary interference with the rest of the school, failure to maintain "discipline," going out of one's way to be argumentative, boasting about better than ordinary results, and other breaches of what is ordinarily thought of as good taste may well promote situations which quickly become intolerable. But to go one's way, quietly and ably and in keeping with scientific-democratic procedure, will go far toward enabling one to choose for himself when he will go to another position.

Available Issues of Kansas Studies in Education

Those issues of *Kansas Studies in Education* which are still in print are listed below. Copies may be secured free of charge by requesting them from the Dean, School of Education, University of Kansas, Lawrence, Kansas.

Vol. 3, No. 1, *A Summary Report to the North Central High Schools of Kansas on Criterion 2: Pupil Activity Program, Library Service, Guidance Service,* by Kenneth E. Anderson. January, 1952, 24 p.

Vol. 3, No. 3, *A Summary Report to the North Central High Schools of Kansas on Criterion 5: School Plant and Equipment,* by Kenneth E. Anderson, with an addendum: "Acoustical Considerations of School Buildings," by James F. Nickerson. April, 1953. 24 p.

Vol. 4, No. 1, *A Summary Report to the North Central High Schools of Kansas on Criterion 3: School Staff,* by Herbert A. Smith and Kenneth E. Anderson. April, 1954. 24 p.

Vol. 4, No. 2, *Body Measurements of Male Students Entering the University of Kansas,* by Edwin R. Elbel. May, 1954. 24 p.

Vol. 4, No. 3, *A Study of 2500 Kansas High School Graduates,* by Kenneth E. Anderson, Carl E. Ladd, and Herbert A. Smith. June, 1954. 45 p.

Vol. 5, No. 1, *Fundamentals for Children in our Time: in Art, Music, Science, Arithmetic, Social Studies, Physical Education, the English Language Arts.* October, 1954. 36 p.

Vol. 5, No. 2, *A Summary Report to the North Central High Schools of Kansas on Criterion 4: Administration and Supervision,* by Herbert A. Smith and Kenneth E. Anderson. April, 1955. 62 p.

Vol. 6, No. 1, *Differentiation of the Phenomenal Field,* by A. H. Turney. July, 1955. 24 p.

Vol. 6, No. 2, *Behavioral Potentials and Kinetics,* by A. H. Turney. February, 1956. 28 p.

Vol. 6, No. 3, *Experiments with Reflective Teaching,* by Ernest E. Bayles. April, 1956. 32 p.